# **A.I** preferences for who to save in self-driving car crashes revealed

Congratulations to: young people, large groups of people, and people who aren’t animals

If self-driving cars become widespread, society will have to grapple with a new burden: the ability to program vehicles with preferences about which lives to prioritize in the event of a crash. Human drivers make these choices instinctively, but algorithms will be able to make them in advance. So will car companies and governments choose to save the old or the young? The many or the few?

A new paper [published today](https://www.nature.com/articles/s41586-018-0637-6) by MIT probes public thought on these questions, collating data from an online quiz [launched in 2016](https://www.theverge.com/2016/8/9/12412190/moral-machine-trolley-problem-self-driving-car) named the Moral Machine. It asked users to make a series of ethical decisions regarding fictional car crashes, similar to the famous [trolley problem](https://en.wikipedia.org/wiki/Trolley_problem). Nine separate factors were tested, including individuals’ preferences for crashing into men versus women, sparing more lives or fewer, killing the young or the elderly, pedestrians or jaywalkers, and even choosing between low-status or high-status individuals.

There are some global preferences, like saving humans over animals

Millions of users from 233 countries and territories took the quiz, making 40 million ethical decisions in total. From this data, the study’s authors found certain consistent global preferences: sparing humans over animals, more lives rather than fewer, and children instead of adults. They suggest these factors should therefore be considered as “building blocks” for policymakers when creating laws for self-driving cars. But the authors stressed that the results of the study are by no means a template for algorithmic decision-making.

“What we are trying to show here is descriptive ethics: peoples’ preferences in ethical decisions,” Edmond Awad, a co-author of the paper, told The Verge. “But when it comes to normative ethics, which is how things should be done, that should be left to experts.”

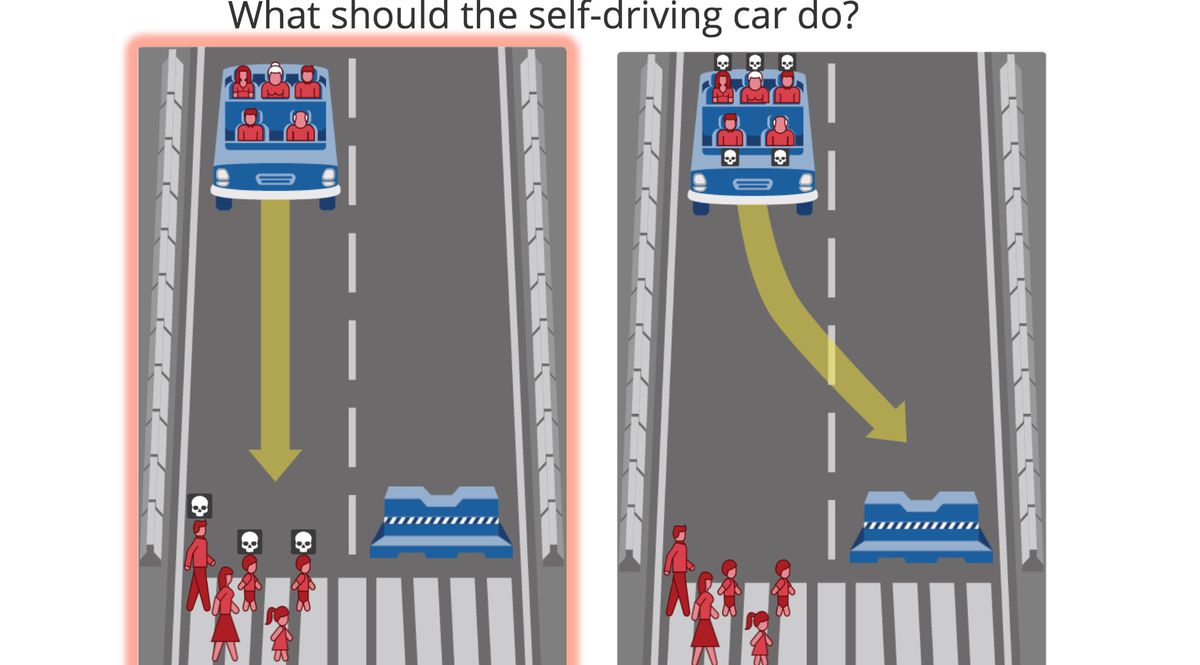
Related

#### [Who will you decide to kill with your self-driving car? Let's find out!](https://www.theverge.com/2016/8/9/12412190/moral-machine-trolley-problem-self-driving-car)

The data also showed significant variations in ethical preferences in different countries. These correlate with a number of factors, including geography (differences between European and Asian nations, for example) and culture (comparing individualistic versus collectivist societies).

It’s important to note that although these decisions will need to be made at some point in the future, self-driving technology still has a way to go. Autonomy is still in its infancy, and self-driving cars (despite [public perception](https://www.theverge.com/2018/10/20/18000884/tesla-full-self-driving-option-gone-musk-autopilot)) are still prototypes, not products. Experts also say that while it’s not clear how these decisions will be programmed into vehicles in the future, there is a clear need for public consultation and debate.

“What happens with autonomous vehicles may set the tone for other AI and robotics, since they’re the first to be integrated into society at scale,” Patrick Lin, director of the Ethics + Emerging Sciences Group at Cal Poly University, told The Verge. “So it’s important that the conversation is as informed as possible, since lives are literally at stake.”

 A sample scenario from the Moral Machine: should the user hit the pedestrians or crash into the barrier?

## How does culture affect ethical preferences?

The results from the Moral Machine suggest there are a few shared principles when it comes to these ethical dilemmas. But the paper’s authors also found variations in preferences that followed certain divides. None of these reversed these core principles (like sparing the many over the few), but they did vary by a degree.

The researchers found that in countries in Asia and the Middle East, for example, like China, Japan, and Saudi Arabia, the preference to spare younger rather than older characters was “much less pronounced.” People from these countries also cared relatively less about sparing high net-worth individuals compared to people who answered from Europe and North America.

Different parts of the world will end up with separate regulatory regimes

The study’s authors suggest this might be because of differences between individualistic and collectivist cultures. In the former, where the distinct value of each individual as an individual is emphasized, there was a “stronger preference for sparing the greater number of characters.” Counter to this, the weaker preference for sparing younger characters might be the result of collectivist cultures, “which emphasize the respect that is due to older members of the community.”

These variations suggest that “geographical and cultural proximity may allow groups of territories to converge on shared preferences for machine ethics,” say the study’s authors.

However, there were other factors that correlated with variations that weren’t necessarily geographic. Less prosperous countries, for example, with a lower gross domestic product (GDP) per capita and weaker civic institutions were less likely to want to crash into jaywalkers rather than people crossing the road legally, “presumably because of their experience of lower rule compliance and weaker punishment of rule deviation.”

The authors stress, though, that the results from the Moral Machine are by no means definitive assessments of different countries’ ethical preferences. For a start, the quiz is self-selecting, only likely to be taken by relatively tech-savvy individuals. It is also structured in a way that removes nuance. Users only have two options with definite outcomes: kill these people or those people. In real life, these decisions are probabilistic, with individuals choosing between outcomes of different severities and degrees. (“If I swerve around this truck, there’s a small chance I’ll hit that pedestrian at a low speed,” and so on.)

Nevertheless, experts say that doesn’t mean such quizzes are irrelevant. The contrived nature of these dilemmas is a “feature, not a bug,” says Lin, because they remove “messy variables to focus in on the particular ones we’re interested in.”

He adds that even if cars won’t regularly have to choose between crashing into object X or object Y, they still have to weigh related decisions, like how wide a berth to give these items. And that is still “fundamentally an ethics problem,” says Lin, “so this is a conversation we need to have right now.”

 A camera peeks through the grille in a Google self-driving car. Photo by Sean O’Kane / The Verge

## Turning ethics into legislation

But how close are we to needing legislation on these issues? When are companies going to start programming ethical decisions into self-driving vehicles?

The short answer to the second question is they already have. This is true in the slightly pedantic sense that every algorithm makes decisions of some sort, and some of those will have ethical consequences. But in more concrete terms, it’s likely that rough preferences are being coded in, even if the companies involved aren’t keen to talk about them publicly.

Companies don’t want to talk about ethical choices

Back in 2014, for example, Google X founder Sebastian Thrun [said](https://www.bloomberg.com/news/videos/2014-10-06/sebastian-thrun-on-evolution-of-selfdriving-car) the company’s prototype self-driving cars would choose to hit the smaller of two objects in the event of a crash. And in 2016, Google’s Chris Urmson [said](http://www.latimes.com/local/transportation/86077633-132.html) its cars would “try hardest to avoid hitting unprotected road users: cyclists and pedestrians.” That same year, a Mercedes-Benz manager reportedly said that the company’s self-driving vehicles would prioritize the lives of passengers in a crash, although the- company later [denied](https://jalopnik.com/now-mercedes-says-its-driverless-cars-wont-run-over-ped-1787890432) this and said it was a misquotation.

It’s understandable that firms aren’t willing to be open about these decisions. On the one hand, self-driving systems are not yet sophisticated enough to differentiate between, say, young and old people. State-of-the-art algorithms and sensors can make obvious distinctions, like between squirrels and cyclists, but they can’t be much more subtle than that. Plus, whichever lives companies might say they prioritize — people or animals, passengers or pedestrians — will be a decision that upsets somebody. That’s why these are ethical dilemmas: there’s no easy answer.

Private companies are doing the most work on these questions, says Andrea Renda, a senior research fellow at the Centre for European Policy Studies. “The private sector is taking action on this, but governments may not find this to be sufficient,” Renda tells The Verge. He says in Europe, the EU is working on ethical guidelines and will likely enforce them through “command and control legislation, or through certification and co-regulation.” In the US, Congress has [published](https://www.wired.com/story/congress-autonomous-self-driving-car-regulations/) bipartisan principles for potential regulation, but federal oversight will likely be slow coming, and it’s not clear whether lawmakers even want to dive into the quagmire of ethical preferences for car crashes.

Renda warns that although the public needs to be involved in these debates, “relying only on bottom-up consultation would be extremely dangerous.” Governments and experts, he says, will need to make choices that reaffirm human rights.

But the problems ahead can already be glimpsed in Germany, the only country to date to propose [official guidelines](https://www.bmvi.de/SharedDocs/DE/Pressemitteilungen/2017/128-dobrindt-massnahmenplan-ethikregeln-fahrcomputer.html) for ethical choices made by autonomous vehicles. Lawmakers tried to slice the Gordian knot of the trolley problem by stating that all human life should be valued equally and that any distinction based on personal features like age or gender should be prohibited. But as the MIT researchers note, if this choice is implemented, it would go against the public’s strong preference for sparing the younger over the elderly. If a government introduces this policy, they ask, how will it handle the backlash “that will inevitably occur the day an autonomous vehicle sacrifices children in a dilemma situation.”

Awad says this sort of conflict is “inevitable” but must be part of the process. “What’s important is to make these decisions transparent,” he says, “to make it clear what’s being done. If this all happens behind the scenes, and people say ‘just trust us,’ I don’t think it will be acceptable. Everyone needs to be involved in these decisions.”